Aalto University
School of Business

Time Series Analysis
Spurious Regression
Unit roots
Cointegration

Tomi Seppälä
Department of Information and Service Management

Spurious Regression

Study the following linear regression model

$$Y_t = \beta_1 + \beta_2 X_t + u_t,\ u_t \sim WN\left(0, \sigma^2\right)$$

where Y_t, X_t are independent Random Walks (unit root processes)
Usually R^2 is high, even though Y_t and X_t are not related in any way.

Why does spurious regression happen? Assume two random walks (RW)

$$Y_t = Y_{t-1} + w_t,\ w_t \sim WN\left(0, \sigma_y^2\right)$$

$$X_t = X_{t-1} + v_t,\ v_t \sim WN\left(0, \sigma_x^2\right)$$

Why does spurious regression happen

Test: $\boldsymbol{H_0}: \beta_{\mathbf{1}} = \beta_{\mathbf{2}} = 0$

against $\boldsymbol{H_1}: \beta_{\mathbf{1}} \neq 0 \ or \ \beta_{\mathbf{2}} \neq 0$

If $\boldsymbol{H_0}$ is true, plugging $\beta_{\mathbf{1}} = \beta_{\mathbf{2}} = 0$ into (SR) gives $\boldsymbol{y_t} = \boldsymbol{u_t}$ is WN but $\boldsymbol{Y_t}$ was supposed to be RW

This is contradiction, since $\boldsymbol{H_0}$ can never be true

→ There is no sense in even testing $\boldsymbol{H_0}$

Note: same argument applies, if you test $\boldsymbol{H_0}: \beta_{\mathbf{2}} = 0$ against $\boldsymbol{H_1}: \beta_{\mathbf{2}} \neq 0$

In this case you can run regression for the differences:

$\Delta \boldsymbol{Y_t} = \boldsymbol{\alpha_1} + \boldsymbol{\alpha_2} \Delta \boldsymbol{X_t} + \boldsymbol{u'_t}$, $\boldsymbol{u'_t} \sim \boldsymbol{WN}\left(\mathbf{0}, \boldsymbol{\sigma^2}\right)$, where $\boldsymbol{\alpha_1} = \mathbf{0}$

If $\boldsymbol{Y_t} \sim \boldsymbol{I}(\mathbf{1})$ and $\boldsymbol{X_t} \sim \boldsymbol{I}(\mathbf{1})$

However, there is no equilibrium solution for this model!

Some properties related to order of integration

1. $X_t \sim I(0) \rightarrow \mathbf{a} + bX_t \sim I(0)$
2. $X_t \sim I(1) \rightarrow \mathbf{a} + bX_t \sim I(1)$
3. $X_t \sim I(0),\ Y_t \sim I(0) \rightarrow \mathbf{a}X_t + bY_t \sim I(0)$
4. $X_t \sim I(1), Y_t \sim I(0) \rightarrow \mathbf{a}X_t + bY_t \sim I(1)$

   X_t dominates, since it is non-stationary

More generally:

$X_t \sim I(d_1), Y_t \sim I(d_2) \rightarrow \mathbf{a}X_t + bY_t \sim I(d)$, where $d = \max(d_1, d_2)$

5. $X_t \sim I(1),\ Y_t \sim J(1)$ then there are two possibilities:

a) $\mathbf{a}X_t + bY_t \sim I(0)$, if $X_t\ and\ \ Y_t$ co-integrated (long-term relationship) or

b) b) $\mathbf{a}X_t + bY_t \sim I(1)$, if $X_t\ and\ \ Y_t$ are not co-integrated

Some properties related to order of integration

Example: $\boldsymbol{X_t = Z_t + u_t}$ $\quad$ $\boldsymbol{u_t \sim WN\left(0, \sigma^2\right)}$ $\quad$ $\boldsymbol{I(0)}$

$$\boldsymbol{Y_t = Z_t + v_t} \qquad \boldsymbol{v_t \sim WN\left(0, \sigma^2\right)} \qquad \boldsymbol{I(0)}$$

If $\boldsymbol{Z_t \sim I(1)}$ then $\boldsymbol{X_t \sim I(1)}$ and $\boldsymbol{Y_t \sim I(1)}$

But $\boldsymbol{X_t - Y_t = u_t - v_t \sim I(0)}$ is stationary

We say that $\boldsymbol{X_t}$ and $\boldsymbol{Y_t}$ are cointegrated of order (1,1):

$\boldsymbol{X_t, Y_t \sim CI(1, 1)}$

Cointegration of order (m,n): CI (m,n)

Definition: Two processes $\boldsymbol{X_t}$ and $\boldsymbol{Y_t}$ are $\boldsymbol{CI}\ (\boldsymbol{m},\boldsymbol{n})$ if

1. $\boldsymbol{X_t} \sim \boldsymbol{I}(\boldsymbol{m})$ and $\boldsymbol{Y_t} \sim \boldsymbol{I}(\boldsymbol{m})$
2. There exist $\boldsymbol{\alpha_1}$ and $\boldsymbol{\alpha_2}$ such that
   $\boldsymbol{\alpha_1 X_t} + \boldsymbol{\alpha_2 Y_t} \sim \boldsymbol{I}(m-n)$

Example: Assume $\boldsymbol{X_t}, \boldsymbol{Y_t} \sim \boldsymbol{I}(1)$

Regression: $\boldsymbol{Y_t} = \beta_1 + \beta_2 \boldsymbol{X_t} + \boldsymbol{u_t}$, $\boldsymbol{u_t} \sim \boldsymbol{WN}\left(\boldsymbol{0}, \boldsymbol{\sigma^2}\right)$

If $\boldsymbol{X_t}$ and $\boldsymbol{Y_t}$ are $\boldsymbol{coin}tegrated\ of\ order\ (1,1)$, i.e. $(\boldsymbol{X_t}, \boldsymbol{Y_t}) \sim \boldsymbol{CI}(1,1)$

then there exist $\boldsymbol{\alpha_1}$ and $\boldsymbol{\alpha_2}$ such that: $\boldsymbol{\alpha_1 X_t} + \boldsymbol{\alpha_2 Y_t} \sim \boldsymbol{I}(0)$

Dividing by $\boldsymbol{\alpha_2}$ does not affect stationarity:

$$\boldsymbol{Y_t} + \frac{\boldsymbol{\alpha_1}}{\boldsymbol{\alpha}_2} \boldsymbol{X_t} \sim \boldsymbol{I}(0)$$

Cointegration of order (m,n): CI (m,n)

Dividing by $\boldsymbol{\alpha_2}$ does not affect stationarity:

$$\boldsymbol{Y_t} + \frac{\boldsymbol{\alpha_1}}{\boldsymbol{\alpha}_2}\boldsymbol{X_t} \sim \boldsymbol{I}(0)$$

Now in the regression equation: $\boldsymbol{u_t} = \boldsymbol{Y_t} - \boldsymbol{\beta_1} - \boldsymbol{\beta_2}\boldsymbol{X_t}$

If we choose $\frac{\alpha_1}{\alpha_2} = -\boldsymbol{\beta_2}$

we get

$$\boldsymbol{u_t} = \boldsymbol{Y_t} - \frac{\alpha_1}{\alpha_2}\boldsymbol{X_t} - \boldsymbol{\beta_1} \sim \boldsymbol{I}(0)$$

Which shows that there exists (Long-run) co-integrating relationship and therefore $\boldsymbol{X_t}$ and $\boldsymbol{Y_t}$ are cointegrated of order (1,1):

Consequences of the order of integration and cointegration on the estimation of linear regression models

$Y_t = \beta_1 + \beta_2 X_t + u_t,\ u_t \sim WN$

Case 1: Both X_t , $Y_t \sim I(0)$ is stationary

Use OLS

Case 2: If $X_t \sim I(d_1), Y_t \sim I(d_2)$, and $d_1 \neq d_2$, i.e. order of integration is different for X_t and Y_t then does *not* make to estimate a regression model. Usually $I(0)$ and $I(1)$

Cointegration

Case 3: If $\boldsymbol{X_t \sim I(d), Y_t \sim I(d)}$, usually ***d = 1***

If same order of integration, two possibilities:

a. $\boldsymbol{X_t}$ and $\boldsymbol{Y_t}$ are not co-integrated

You can (and have to) use a difference model (of order d)

E.g. $\Delta \boldsymbol{Y_t} = \boldsymbol{\beta_1} \Delta \boldsymbol{X_t} + \boldsymbol{u_t}$ (*but there is no LT relationship*)

Note: There is no constant term

b. If $\boldsymbol{X_t}$ and $\boldsymbol{Y_t} \sim \boldsymbol{I(d,d)}$; usually ***d = 1*** then possibly we could use difference model, but much better to use, and you should use the

Error Correction Model (ECM): If ***d = 1***

$$\Delta \boldsymbol{Y_t} = \boldsymbol{\beta_1} \Delta \boldsymbol{X_t} + \boldsymbol{\beta_2} (\boldsymbol{X_{t-1}} - \gamma \boldsymbol{Y_{t-1}}) + \boldsymbol{u_t}$$

where $\boldsymbol{X_t} - \gamma \boldsymbol{Y_t} \sim \boldsymbol{I(0)}$

Error Correction term (Co-integrating relationship)